Medieval Citation Networks as Digital Hyperlinks:

AI-Powered Authorship Attribution for Historical Digital Libraries

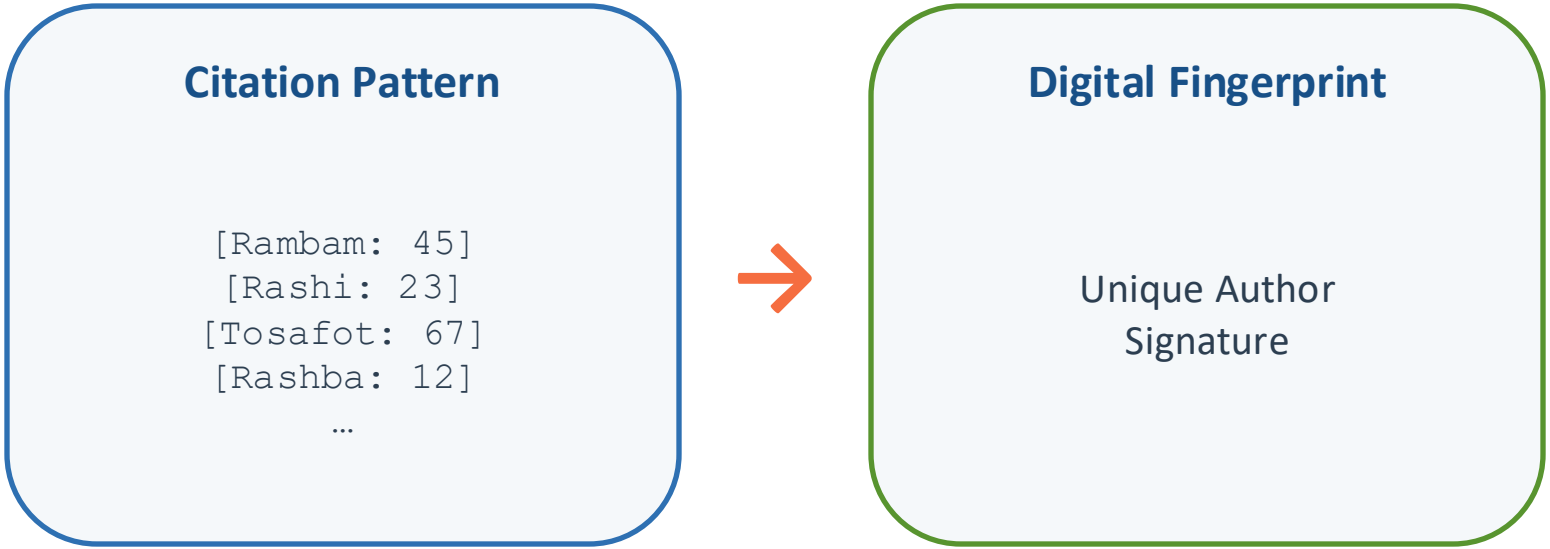
Jonathan Schler, Nati Ben-Gigi, Binyamin Katzoff, Maayan Geffet-Tamir

WADL 2025 | HT Conference, Chicago
Sep 15th, 2025

מכון טכנולוגי חולון
Holon Institute of Technology

Bar-Ilan University
אוניברסיטת בר-אילן

Can Citation (HT) Patterns Serve as Digital Fingerprints?



Can this scale to millions of documents?

Why Digital Libraries Need AI-Powered Attribution

- Historical manuscripts often lack reliable authorship metadata
  - Manual attribution doesn't scale to mass digitization
- Medieval texts: implicit citation networks locked in unstructured text
  - Citations are ancient 'hyperlinks' - navigable knowledge waiting to be unlocked
  - No existing database of these reference networks

Challenge: Extract networks → Create fingerprints → Resolve attribution

Medieval Citations = Ancient Hyperlinks

Medieval Citation

"As the Rambam writes in Hilkhot Teshuvah chapter 3…"

- Creates pathway
- Encodes authority
- Locked in prose

HTML Hyperlink

*For further explanations see Maimonides on Repentance *

- Creates pathway
- Encodes authority
- Machine-readable

Both create navigable knowledge networks

Terumat HaDeshen, sub-section 2

Question:

If someone is reciting the Shema and the prayer leader (Shaliach Tzibur) reaches Modim while they are in the middle of a chapter or a blessing…

Answer:

It appears that there is evidence that one may interrupt and respond to Modim. In the chapter One Whose Deceased Lies Before Him, Rav Huna says: "If someone enters the synagogue and finds the congregation praying, if they can finish their prayer before the prayer leader reaches Modim, they should pray, and if not, they should not pray." From here, we see that the Talmud is strict to the extent that one should delay their prayer to respond to Modim, and we do not say they should pray and bow without responding in the middle of the blessing. As Tosfotis and the Rosh wrote in the name of Rabbeinu Tam, it was his custom to do so when he was praying and the prayer leader reached Modim…We do not find in any source that these *Amens* are superior to other *Amens*, except for what the Rosh brings in the Jerusalem Talmud in the chapter One Whose Deceased Lies Before Him: "If someone enters the synagogue

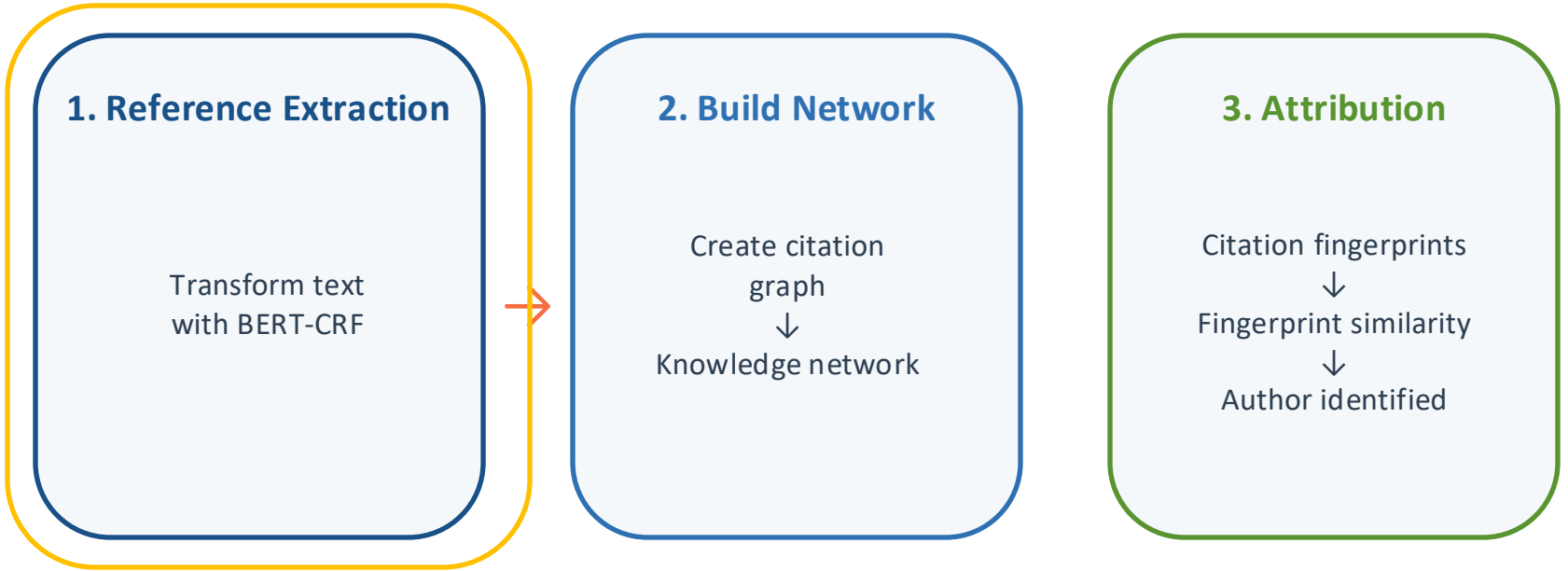
A reference to a chapter in the Talmud with no details of the specific tractate

Recursive reference

Consecutive references

A reference with hierarchical components

From Unstructured Text to Knowledge Graphs



Bridge medieval scholarship with modern digital library infrastructure

Unlocking References from Historical Text

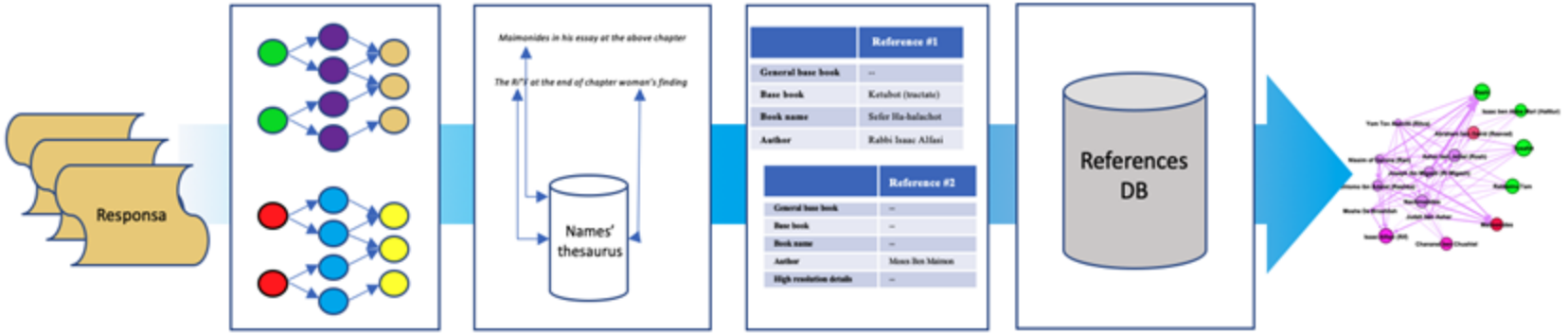
Three-stage classification process:

- Stage 1: Reference boundary detection
  - Identify where citations begin and end
  - Handle consecutive and nested references
- Stage 2: Component identification
  - Author names, book titles, chapter/page references
- Stage 3: Entity normalization
  - Author names, book titles, chapter/page canonization
  - Handle abbreviations and spelling variations

Training: 3,301 references | 20,477 entities | F1 = 0.901

Multi-Layer Reference Identification System

Design & Components



Responsa corpus

Internal component and reference boundary identification

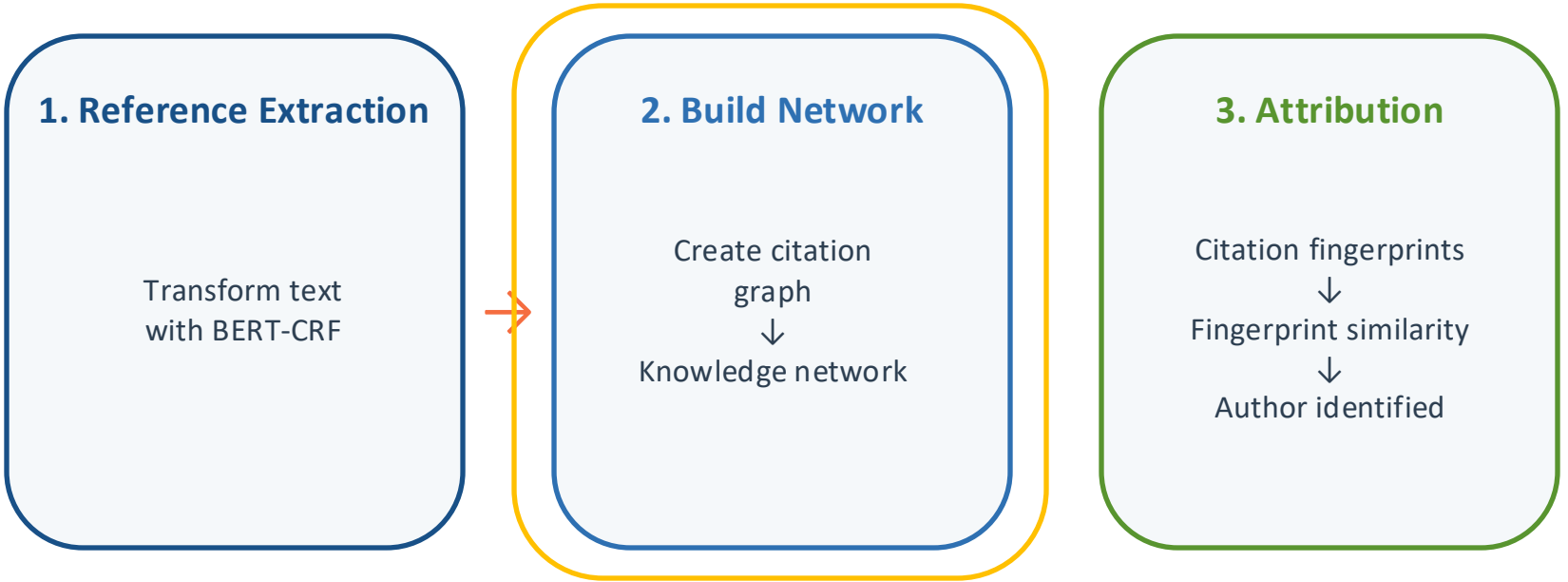
Entities name formalization using names' thesaurus

References finalization and validation. Completion of missing details

References DB construction

Final network construction

From Unstructured Text to Knowledge Graphs



Bridge medieval scholarship with modern digital library infrastructure

Building a Medieval Knowledge Graph

62.5M

Tokens
Processed

230K+

Citations
Extracted

250

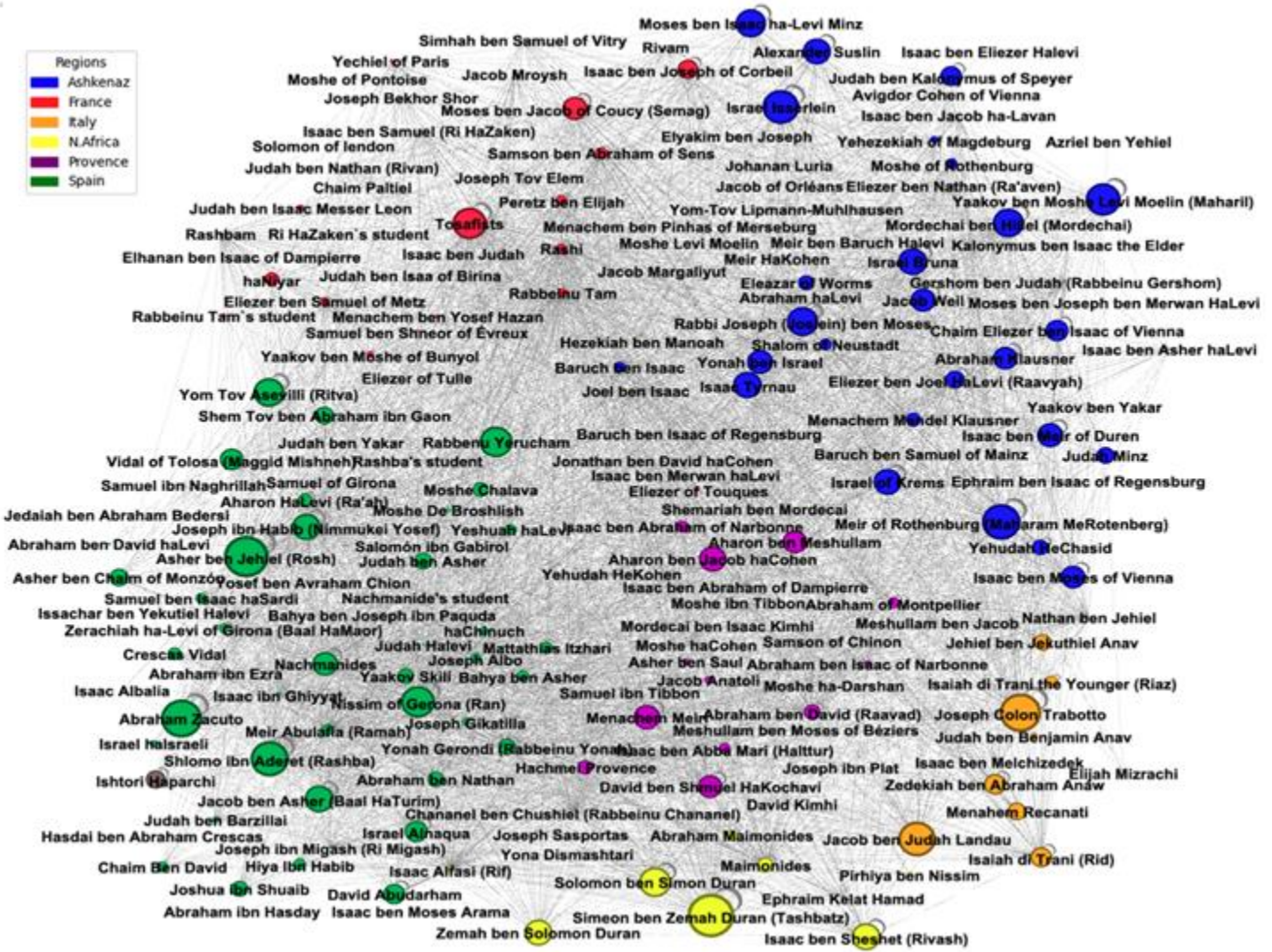
Books
From 124 Authors
Analyzed

6

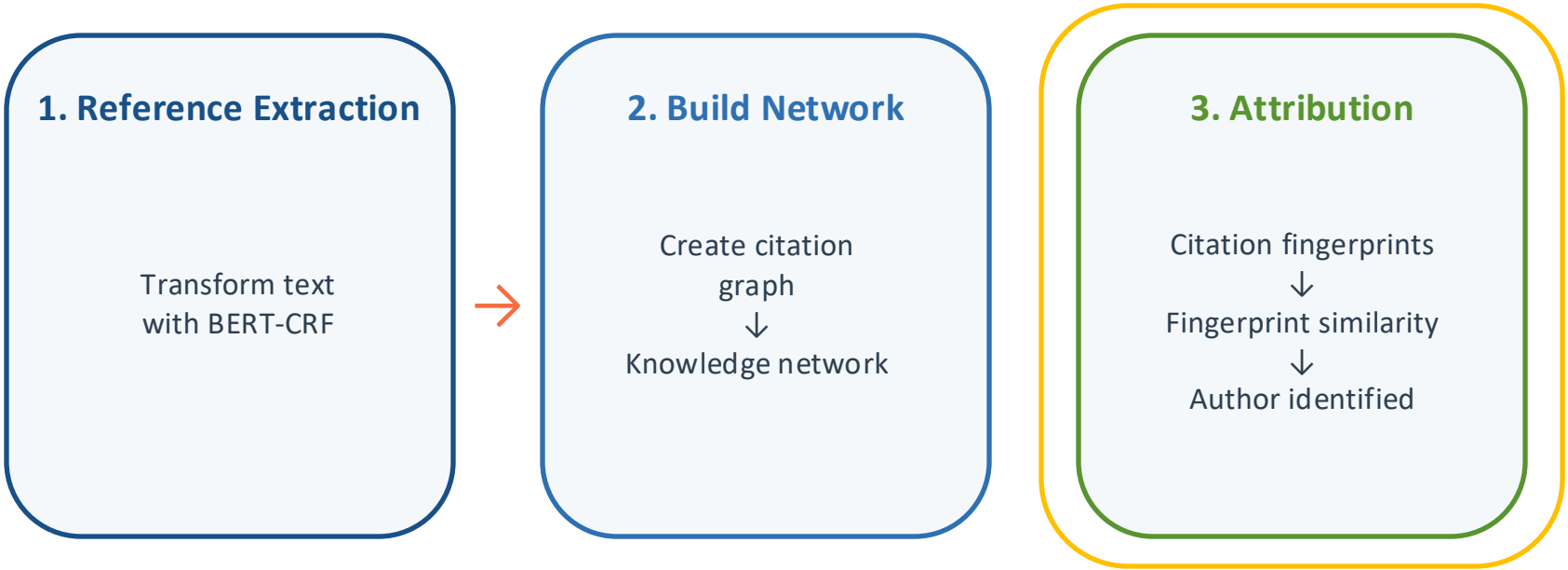
Geographic
Regions

First comprehensive citation network for medieval Hebrew literature

The Entire Corpus Citation Network



From Unstructured Text to Knowledge Graphs



Bridge medieval scholarship with modern digital library infrastructure

From Networks to Signatures

Mathematical Representation:

$$v\_author = [c_1, c_2, \ldots, c_n]$$

where c_i = frequency of citing authority i

Comparison Method:

$$\text{similarity}(\mathbf{v}_a, \mathbf{v}_b) = \frac{\mathbf{v}_a \cdot \mathbf{v}_b}{|\mathbf{v}_a| \cdot |\mathbf{v}_b|}$$

- Language-agnostic approach
- Works regardless of text length
- Range: 0 (different) to 1 (identical)

Do Citation Patterns Remain Stable?

Known Authors	Consistency Score
Ramban	**0.99**
Rashba	**0.98**
Ritva	**0.98**
Maharam Chalava	**0.98**

✓ Citation patterns are reliable authorial signatures

Stable across different works by the same author

The Ritva Attribution Problem

A 300-year scholarly debate

- Commentary on Tractate Bava Metzia
- Two texts circulate under Ritva's name:
  - Hiddushei HaRitva (accepted as authentic)
  - Amsterdam 1729 edition (disputed)
- 18th century scholars suspected multiple authors
- Traditional methods remain inconclusive

Perfect test case for our computational approach

Computational Approach Applied

- **Step 1:** Citation extraction (847 citations from 1729 edition, sufficient statistically (>530))

- **Step 2:** Baseline comparison:
  Authentic Ritva works to itself (all accepted commentaries) (μ = 0.82 , σ = 0.11)

- **Step 3:** Discovered structural break:
  Initial analysis: low similarity to Ritva (0.32)
  Hypothesis: Check for multiple authors
  Split at folio 12 based on scholarly suggestions

Manuscript Structure

Part 1: Folios 1-12

Similarity to Ritva:

0.87

Best match:

Kreskes Vidal (0.91)

Student of Rashba

Part 2: Folios 12-end

Similarity to Ritva:

0.17

Best match:

Ibn Gaon (0.959)

Previously unknown
Student of Rashba

Confirms scholarly suspicion of multiple authors

Computational Evidence

Author	Similarity Score	Interpretation
R. Shem Tov ibn Gaon	**0.959**	Match!
Ra'avad the third	**0.835**	Possible
R. Shmuel Hasardi	**0.728**	Unlikely
Authenticated Ritva	**0.170**	Ruled out

Clear quantitative answer to centuries-old question

Identifying Rabbi Shem Tov ibn Gaon

c. 1250-1330, Soria, Northern Castile

- Student of the Rashba
- Author of Migdal Oz (commentary on Maimonides)
- Explicitly mentioned writing a "Shita" on Bava Metzia
- This work was considered lost
- Similarity score: 0.959 (highest in entire corpus)

First computational identification of this attribution

Scaling Beyond Medieval Hebrew

Methodology applies to any citation-rich corpus:

- Classical Greek and Latin texts
  - Philosophical works, historical chronicles
- Arabic manuscript traditions
  - Islamic scholarship, scientific treatises
- Early modern scientific literature
  - Royal Society papers, correspondence networks
- Legal precedent networks
  - Case law citations, judicial opinions
- Modern academic papers
  - ArXiv, PubMed, institutional repositories

Beyond Attribution: Infrastructure for Scholarship

- Created navigable hypertext from static manuscripts
  - 230,000+ extracted links between texts
- Knowledge graphs from citation networks
  - Reveals influence patterns and scholarly communities
- Scalable to any citation-rich corpus
  - Language-agnostic methodology
- Enables new forms of scholarly exploration
  - Navigate medieval texts like modern web

Transform historical collections into interactive digital libraries

Ancient Hypertext Meets Modern AI

- Medieval citations were the first hyperlinks
  - Created navigable knowledge networks centuries ago
- AI can unlock these networks at scale
  - Transformer models extract hidden structure
- Citation fingerprints provide reliable attribution
  - Quantitative solution to scholarly debates
- Digital libraries gain both navigation and verification
  - Infrastructure for 21st century scholarship

Bridging ancient wisdom with modern technology

Thank You

Questions?

Jonathan Schler

schler@hit.ac.il